



Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D.,  
 Dickson M.C., Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
 Butcherfield V.S., Krzywinski M.I., Skalska U., Smallwood D.E.,  
 Scherch A., Schein J.E., Jones S.J., and Marra M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL  
 PUBLISHED 12477932  
 REFERENCE 2 (bases 1 to 3248)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Lohar Hemmighausen Ph.D., Robin Humphreys  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdpaxil.stanford.edu](mailto:mdpaxil.stanford.edu)  
 Dickson, M., Schmitt, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAC Plate: 7 Row: h Column: 16  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6677898  
 This clone has the following problem: no 5' EST match.

FEATURES  
 source location/qualifiers

1..3248  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3489914"  
 /tissue\_type="Mammary tumor. WAP-TGF alpha model. 7 months  
 old. gross tissue."  
 /clone\_id="NCI CGAP\_Mam5"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 99.9%; Score 3102.8; DB 11; Length 3248;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGGTAGACGGCTGCCCCGAGAGGACACGCGTCTGAGACCGGCGATGGAACCGCCAAA 60  
 DB 111 CATGGTAGACGGCTGCCCCGAGAGGACACGCGTCTGAGACCGGCGATGGAACCGCCAAA 170  
 QY 61 CCAATGGTAGCAATGCGGGCGCCGCAAGCGCGGAGCTTCAGGACTTCGCGCGGAGAC 120  
 DB 171 CCAATGGTAGCAATGCGGGCGCCGCAAGCGCGGAGCTTCAGGACTTCGCGCGGAGAC 230  
 QY 121 TCAAGTACCACTCCCGGCTAGAGAACATGAATGGCTTTGAGAGGGGTGTGAGTTCTGTC 180  
 DB 231 TCAAGTACCACTCCCGGCTAGAGAACATGAATGGCTTTGAGAGGGGTGTGAGTTCTGTC 290  
 QY 181 CTGCGAACATGCGCAAGAAAGTGAAGAGCGAGGCCCGACGCGCTGGGTGTGTGTGTG 240  
 DB 291 CTGCGAACATGCGCAAGAAAGTGAAGAGCGAGGCCCGACGCGCTGGGTGTGTGTGTG 350  
 QY 241 CAGTGTCTTCAAGCTTCTCTGCTCTCTCTCAATGAGCTGGCTGGTGTGTGTGTGCACTTCC 300  
 DB 351 CAGTGTCTTCAAGCTTCTCTGCTCTCTCTCTCAATGAGCTGGCTGGTGTGTGTGTGCACTTCC 410  
 QY 301 ATTATCGAATGTGGGGTTCAAAAAGTCTTCATATGAGCTGAGATCAACAAATGAGA 360

DB 411 ATTATCGAATGTGGGGTTCAAAAAGTCTTCATATGAGCTGAGATCAACAAATGAGA 470  
 QY 361 TCTTTCTGATGCGATATGAGAACTCACTCCACAGATTATACAGCTGGCCAGCAG 420  
 DB 471 TCTTTCTGATGCGATATGAGAACTCACTCCACAGATTATACAGCTGGCCAGCAG 530  
 QY 421 TGAAGAGGCGCTGAGAGCTGCTGTACATGAATGCTCTGCTGGTCCCTACCAAGA 480  
 DB 531 TGAAGAGGCGCTGAGAGCTGCTGTACATGAATGCTCTGCTGGTCCCTACCAAGA 590  
 QY 481 AGTGGCTGTAACTGCTTCACTGAGAGGAGAGTCAATGCGCTACTGCTGAGATTCA 540  
 DB 591 AGTGGCTGTAACTGCTTCACTGAGAGGAGAGTCAATGCGCTACTGCTGAGATTCA 650  
 QY 541 GCATCCCCCAACCTGGGAGAGAGAGTGTATGCGCCCATAGCTGTGAGCGAGTTGAA 600  
 DB 651 GCATCCCCCAACCTGGGAGAGAGAGTGTATGCGCCCATAGCTGTGAGCGAGTTGAA 710  
 QY 601 CATTCGCAACCCCGAGACCGGCACTGAATCTCTTCTGTCTTAACATCTGTGAGCTTCC 660  
 DB 711 CATTCGCAACCCCGAGACCGGCACTGAATCTCTTCTGTCTTAACATCTGTGAGCTTCC 770  
 QY 661 CCAATTGACCCCAATGCTGTGAGAGAGCTCAGAGCAACAGTGTGAGTTGGCTGCATG 720  
 DB 771 CCAATTGACCCCAATGCTGTGAGAGAGCTCAGAGCAACAGTGTGAGTTGGCTGCATG 830  
 QY 721 CCCATGGTGAGAGAGTACAGCTTCACTACCCCTGCTTCCCAACAGTCTCTACCCG 780  
 DB 831 CCCATGGTGAGAGAGTACAGCTTCACTACCCCTGCTTCCCAACAGTCTCTACCCG 890  
 QY 781 CGCATGCGCGCTGCGATGCTGCTGCGGGGAGACGCTGCTGTGTGTGTGTGTGTGTGT 840  
 DB 891 CGCATGCGCGCTGCGATGCTGCTGCGGGGAGACGCTGCTGTGTGTGTGTGTGTGTGT 950  
 QY 841 TCCGAAGCTTTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 DB 951 TCCGAAGCTTTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1010  
 QY 901 ATAGCTGAGACCCCGATGAGAACCCCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 960  
 DB 1011 ATAGCTGAGACCCCGATGAGAACCCCAACGCTGTGTGTGTGTGTGTGTGTGTGTGT 1070  
 QY 961 CTTAACAACCTGACTTCT 1020  
 DB 1071 CTTAACAACCTGACTTCT 1130  
 QY 1021 CTGACCGGCGACATCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080  
 DB 1131 CTGACCGGCGACATCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1190  
 QY 1081 GTGCGGCTTTTGTAGTGACACCCCAAGGAGCATTTAGCAGGCTCTACTATTCAGGCCACT 1140  
 DB 1191 GTGCGGCTTTTGTAGTGACACCCCAAGGAGCATTTAGCAGGCTCTACTATTCAGGCCACT 1250  
 QY 1141 ACCGCGCCCAATCACTGACATGCAATGAAATATCAAGGTGCCCAACACCGGAACGTAAG 1200  
 DB 1251 ACCGCGCCCAATCACTGACATGCAATGAAATATCAAGGTGCCCAACACCGGAACGTAAG 1310  
 QY 1201 TGGCTTCAACCTCTTCTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260  
 DB 1311 TGGCTTCAACCTCTTCTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1370  
 QY 1261 ACTATGTGAGATCAACCGGAGAGAGTACTGCGGTGAGAGGTCCCAATTTGTGTGTGTG 1320  
 DB 1371 ACTATGTGAGATCAACCGGAGAGAGTACTGCGGTGAGAGGTCCCAATTTGTGTGTGTG 1430  
 QY 1321 GCAACAGACAGAAATTAAGTTCAGTTCATCTTCTGATCACTGATCAACCGGAGT 1380  
 DB 1431 GCAACAGACAGAAATTAAGTTCAGTTCATCTTCTGATCACTGATCAACCGGAGT 1490  
 QY 1381 TCCCTGTGAGATCACTTCTCTTCAAGCTCCCAAGACCCGCTCCAGAGATGTTCAATGTGA 1440

Db 1491 TCCTAGCTGAGTACCTCTCTACGACTCAACGACCCGCTGCCAGGATGTTTCAATGCA 1550

QY 1441 AGACTGACGGTGCATCCGAAAGAAATCTGGCTGCGACGCGCTGGGAGACTGCCGGATT 1500

Db 1551 AGACTGACGGTGCATCCGAAAGAAATCTGGCTGCGACGCGCTGGGAGACTGCCGGATT 1510

QY 1501 ATAGTGAATGAGCGTTACTGCGAATGCAATGCCACCCGATTCACTGCAAAAACCACT 1560

Db 1611 ATAGTGAATGAGCGTTACTGCGAATGCAATGCCACCCGATTCACTGCAAAAACCACT 1670

QY 1561 TCTGCAAGCCCTCTTCTGGGCTCTGTCAGTGTCAAGCTGTGGGGACGGAAGTAGC 1620

Db 1671 TCTGCAAGCCCTCTTCTGGGCTCTGTCAGTGTCAAGCTGTGGGGACGGAAGTAGC 1730

QY 1621 AGAGAGGCTGACGCTGCTCTGTCGGAGTTTCAAGTGTTCATGGAAGTGTCTCCCTC 1680

Db 1731 AGAGAGGCTGACGCTGCTCTGTCGGAGTTTCAAGTGTTCATGGAAGTGTCTCCCTC 1790

QY 1681 AGAGCCAGAGTGTATGGAAGAGCACTGTGGAATGGGTCTGACGAGCTTCAATGTC 1740

Db 1791 AGAGCCAGAGTGTATGGAAGAGCACTGTGGAATGGGTCTGACGAGCTTCAATGTC 1850

QY 1741 ACAGCGTGAATGCTGCTCTTGGACCAAAATATCTACCGCTGCCAAAATGGCTCTGTC 1800

Db 1851 ACAGCGTGAATGCTGCTCTTGGACCAAAATATCTACCGCTGCCAAAATGGCTCTGTC 1910

QY 1801 TGAGCAGAGGCAACCTTGATGTGATGGAAGAAGAGCTGTAGCGATGAGCTCCGATGAGA 1860

Db 1911 TGAGCAGAGGCAACCTTGATGTGATGGAAGAAGAGCTGTAGCGATGAGCTCCGATGAGA 1970

QY 1861 AAAAAGTGAATGCTGCTGCTGCTGCTTACCAAAAGAGCTGGCGTGTGGTGGAGCA 1920

Db 1971 AAAAAGTGAATGCTGCTGCTGCTGCTTACCAAAAGAGCTGGCGTGTGGTGGAGCA 2030

QY 1921 ATGCGAGCAGGAGGAGTGGCCCTGCGACAGTGAAGCTCCAGCCCTGGGCGAGGCGACT 1980

Db 2031 ATGCGAGCAGGAGGAGTGGCCCTGCGACAGTGAAGCTCCAGCCCTGGGCGAGGCGACT 2090

QY 1981 TGTGTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040

Db 2091 TGTGTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2150

QY 2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGGAAGTGGGCTTCTGGGCTGCTGCTG 2100

Db 2151 ATGACAAAATTTCAAGTACTCAGACTACAGATGGAAGTGGGCTTCTGGGCTGCTGCTG 2210

QY 2101 ACCAGAGCAGGAGGAGTGGCTCTGCGGCTGCGAGAGTGAAGCTCAAAAGTATCATCAACC 2160

Db 2211 ACCAGAGCAGGAGGAGTGGCTCTGCGGCTGCGAGAGTGAAGCTCAAAAGTATCATCAACC 2270

QY 2161 ACCCTTCTTCAATGATTTTCACTTTCGATGATGATGATGATGATGATGATGATGATGAT 2220

Db 2271 ACCCTTCTTCAATGATTTTCACTTTCGATGATGATGATGATGATGATGATGATGATGAT 2330

QY 2221 CGGTGAGTGAAGCAGCGTGGGCGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280

Db 2331 CGGTGAGTGAAGCAGCGTGGGCGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2390

QY 2281 CTGCTGGCAGGCGCATCTGAGGTCAAGGCTGGGCGCACAAAAGAGGAGTACCGGAG 2340

Db 2391 CTGCTGGCAGGCGCATCTGAGGTCAAGGCTGGGCGCACAAAAGAGGAGTACCGGAG 2450

QY 2341 CGCTGATCTTGCAGAAAGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400

Db 2451 CGCTGATCTTGCAGAAAGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2510

QY 2401 TGCGCAGCAGATCAACCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460

Db 2511 TGCGCAGCAGATCAACCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2570

QY 2461 CTGCGCAGGCTGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520

Db 2571 CTGCGCAGGCTGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2630

QY 2521 AGGCTGATGTGATGAGCTGGGCTGAAGGCTGCCTCAGAGAGAAACAAGCCAGGCTGTACA 2580

Db 2631 AGGCTGATGTGATGAGCTGGGCTGAAGGCTGCCTCAGAGAGAAACAAGCCAGGCTGTACA 2590

QY 2581 CAAGGCTCCTCTGATGTTGGGACTGAGTCAAGAGCAACTGGGCTATATGACAGATGAC 2640

Db 2691 CAAGGCTCCTCTGATGTTGGGACTGAGTCAAGAGCAACTGGGCTATATGACAGATGAC 2750

QY 2641 AGACAGCCGACCAACAACACCCAGAGATGCCGACATGACACCTGATACAGAGAG 2700

Db 2751 AGACAGCCGACCAACAACACCCAGAGATGCCGACATGACACCTGATACAGAGAG 2810

QY 2701 GAACACTGACGACTTTATGCTGTGGCTCCCCCCCCCAACAACAACCCGACTGTGAACT 2760

Db 2811 GAACACTGACGACTTTATGCTGTGGCTCCCCCCCCCAACAACAACCCGACTGTGAACT 2870

QY 2761 GCATCCTTAGAATCAAGATTTCTTCCAAAGTGGGACCCCTCAAGAGTTGAGAGAGAAC 2820

Db 2871 GCATCCTTAGAATCAAGATTTCTTCCAAAGTGGGACCCCTCAAGAGTTGAGAGAGAAC 2930

QY 2821 TTGCGCTAGAGGCGCCACCTGGGCGCAAGGCTTTGATGAGAGCTTCCCTCTAGCC 2880

Db 2931 TTGCGCTAGAGGCGCCACCTGGGCGCAAGGCTTTGATGAGAGCTTCCCTCTAGCC 2990

QY 2881 CTGAGCTGGTGAAGATGATGCTGCTCCGAGAGGCTTCCCACTGATGAGGCTCCC 2940

Db 2991 CTGAGCTGGTGAAGATGATGCTGCTCCGAGAGGCTTCCCACTGATGAGGCTCCC 3050

QY 2941 GGGAGCCTATGAGGAGAGGCTCAAGGCTCACTTTTCAGAGAGCGCAGCCTTAGA 3000

Db 3051 GGGAGCCTATGAGGAGAGGCTCAAGGCTCACTTTTCAGAGAGCGCAGCCTTAGA 3110

QY 3001 ACCCGAGAAAGAGTGAATCTTAAGCTGAATGTTGTTGCTGCTTCCAGGCTGTAT 3060

Db 3111 ACCCGAGAAAGAGTGAATCTTAAGCTGAATGTTGTTGCTGCTTCCAGGCTGTAT 3170

QY 3061 TTGAGAGTAAACATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 3106

Db 3171 TTGAGAGTAAACATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 3216

RESULT 2  
AK052738  
LOCUS  
DEFINITION  
Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone: D63041B06 product: suppression of tumorigenicity 14 (colon carcinoma), full insert sequence.  
AK052738  
ACCESSION  
AK052738.1 GI:26342936  
VERSION  
HNC; CAP trapper.  
KEYWORDS  
Mus musculus (house mouse)  
SOURCE  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, O., Nishi, K., Kitanishi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6  
(bases 1 to 4017)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirokane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/.  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers  
1. 4017  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:D630041B06"  
/db\_xref="MG1:2432676"  
/db\_xref="MG1:10090"  
/clone="D630041B06"  
/tissue\_type="Kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
174..2741  
/note="unnamed protein product, putative suppression of tumorigenicity 14 (colon carcinoma) [MGI:MGI:1338881, GB|NM\_011176, evidence: BLASTN, 100%, match=720]"  
/codon\_start=1  
/protein\_id="AAC35125.1"  
/db\_xref="GI:26342937"  
/translation="MGSRGRKAGGGGSGDPSAGLKYNSRLNNNGFEGVEFLPANNAKVKEGRRWVVAIVLPSFLILSLAGLLVWHPRYNNVVOVKFENHLITIEFLDAVNSTEFLSLASGVKALKILYNVPVLPYHKKSAVTAESGSLVIAVYSEFSIPHLAEVDAAVMAVERVTLPPRARALKSPVLSVAFPIDPRLQGTQNSCSFAL

## ORIGIN

Query Match 99.4%; Score 3087.6; DB 11; Length 4017;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3090; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

HAHGAATRTTTCGFPNSFPYAHARCOVLRGDADSLVLSIFRSPDVAFCDEHSDLV  
TYVDSLSPMEPAVVRLCSTFSPSYNLTLSSQNVFLVTLINTRRHGFPAFFOL  
TPMSSCGGLSTOGTGESPPYRGXPYPINCCNMVKNVNNVKKRFLAFLVBNV  
PYGSCTRDYVEINRGKTCYCBRSOPVSSKITYHSDHFTDNGKFOCKPLFWVC  
DECPMFWKTCRCIRKELRCGMADCPYSPBRICRCAATHTQCTNFOCKPLFWVC  
DSVNOCGDSDDEGSCCPAGSFRCSNCKLPSSQKNCNDKNSFTQARVAGTNDG  
CTKYVRCNGILCSKGNDECDKTDSCDSGDEKDCNRKYSYDVTQARVAGTNDG  
EWPQVSHALGQHLGASLISPDVLSAAGCPQDKNFKYSDYTKVAFGLDQ  
KRSAGVQELKRIITTHPSFNDFTPDVIALLEKSEYESTVVRPILDPATVFP  
AKRALVYTWGVTGRTGALLQKGEIRVINGTCEIDMPOIIPRMVCGFLSGV  
DSCQDSSGFLSSAEKXDRMFPQAGVYSSWEGCAQNRKRPVYTRLPVVRMIENHGV

QY 1 CATGTAGACGGCTGCCGAGGAGCCAGCGCTTGAAGACCGGCAATCGACCGCCAAA 60  
DB 112 CATGTAGACGGCTGCCGAGGAGCCAGCGCTTGAAGACCGGCAATCGACCGCCAAA 171  
QY 61 CCATGGGTGCAATCGGGGCGCCAGGCGGAGGGGCTCTCAGGACTTCGCGCGGAGAC 120  
DB 172 CCATGGGTGCAATCGGGGCGCCAGGCGGAGGGGCTCTCAGGACTTCGCGCGGAGAC 231  
QY 121 TCAAGTACAACCTCCGCGCTAGAGACATGATGCTTTGAGAGGGTGTGAGTTCCTGC 180  
DB 232 TCAAGTACAACCTCCGCGCTAGAGACATGATGCTTTGAGAGGGTGTGAGTTCCTGC 291  
QY 181 CTCGGAACAATGCCAAGAAAGTGAAGAGGAGCGCCACGCGCTGGTGTGCTGTGG 240  
DB 292 CTCGGAACAATGCCAAGAAAGTGAAGAGGAGCGCCACGCGCTGGTGTGCTGTGG 351  
QY 241 CAGTGGTTCACGCTTCCTGCTGCTCCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
DB 352 CAGTGGTTCACGCTTCCTGCTGCTCCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 411  
QY 301 ATTATCGGAATGTGCGGGTCAAAAAGCTCTTCAATGCGCATCTGAGATCAAAATGAGA 360  
DB 412 ATTATCGGAATGTGCGGGTCAAAAAGCTCTTCAATGCGCATCTGAGATCAAAATGAGA 471  
QY 361 TCTTTCTGATGGGTATGAGAACTCCACCTCCACAGAGTTTATGAGGCGCGGAGG 420  
DB 472 TCTTTCTGATGGGTATGAGAACTCCACCTCCACAGAGTTTATGAGGCGCGGAGG 531  
QY 421 TGAAGAGGGGCTGAGACCTGCTGTAATGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
DB 532 TGAAGAGGGGCTGAGACCTGCTGTAATGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 591  
QY 481 AGTCGGCTGTAACCTGCTTCAGTAGAGGCAAGTGTCAATGCTCTACTGCTGAGAGTTCA 540  
DB 592 AGTCGGCTGTAACCTGCTTCAGTAGAGGCAAGTGTCAATGCTCTACTGCTGAGAGTTCA 651  
QY 541 GCATCCCCCAGACCTGCGAGAAAGTGTATGCGGCGCATGCGGTGAGGAGGAGTTGTA 600  
DB 652 GCATCCCCCAGACCTGCGAGAAAGTGTATGCGGCGCATGCGGTGAGGAGGAGTTGTA 711  
QY 601 CATGCGCACCCGAGACGAGGCACTGAATCTCTGCTGTAACATCTGTGTGAGCTTCC 660  
DB 712 CATGCGCACCCGAGACGAGGCACTGAATCTCTGCTGTAACATCTGTGTGAGCTTCC 771  
QY 661 CCATGACCCGAGATGCTGAGAGAGACTGAGACACAGTGTGAGTTTGGCTGAGT 720  
DB 772 CCATGACCCGAGATGCTGAGAGAGACTGAGACACAGTGTGAGTTTGGCTGAGT 831  
QY 721 CCATGAGTGCAGAGTGAACAGCTTCACTACCCCTGCTCCCAACAGTCCCTACCGG 780  
DB 832 CCATGAGTGCAGAGTGAACAGCTTCACTACCCCTGCTCCCAACAGTCCCTACCGG 891  
QY 781 CGCATGCGCGGCGCAGTGGGTCTGCGCGGGAGGACGCGGACTCTGTGCTGAGCTTCACT 840  
DB 892 CGCATGCGCGGCGCAGTGGGTCTGCGCGGGAGGACGCGGACTCTGTGCTGAGCTTCACT 951

QY 841 TCGAAGCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAACCTGGTCAACCGTGTATG 900  
DB 952 TCGGAACCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAACCTGGTCAACCGTGTATG 1011  
QY 901 ATAGCTGAGCCCCATGGAACCCCAAGCTGTGTGCGGCTGTGTGACCTTTCTCAACCT 960  
DB 1012 ATAGCTGAGCCCCATGGAACCCCAAGCTGTGTGCGGCTGTGTGACCTTTCTCAACCT 1071  
QY 961 CCTACAACTGACTTTCTCTCTCTCCCAAGACGTCTTCTGTGTACGCTGTAAACCAATA 1020  
DB 1072 CCTACAACTGACTTTCTCTCTCCCAAGACGTCTTCTGTGTACGCTGTAAACCAATA 1131  
QY 1021 CTGACCGGCGACATCCGCTTTGAGGCCACTTTCTCCAGCTGCCCAAGATAGACGCT 1080  
DB 1132 CTGACCGGCGACATCCGCTTTGAGGCCACTTTCTCCAGCTGCCCAAGATAGACGCT 1191  
QY 1081 GTGGCGGCTTTTGTAGTGAACCCCAAGGACATTTAGACGCCCTCTCTATCCAGGCGACT 1140  
DB 1192 GTGGCGGCTTTTGTAGTGAACCCCAAGGACATTTAGACGCCCTCTCTATCCAGGCGACT 1251  
QY 1141 ACCCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCAACACCGGAACGTGAAG 1200  
DB 1252 ACCCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCAACACCGGAACGTGAAG 1311  
QY 1201 TGGGCTTCAAACTCTTCTATCTGTGTGAGGCCCAAGCTGACGCGGCTCTGTACCAAG 1260  
DB 1312 TGGGCTTCAAACTCTTCTATCTGTGTGAGGCCCAAGCTGACGCGGCTCTGTACCAAG 1371  
QY 1261 ACTATGTGAGATCAACCGGAGAAAGTACTGCGGTGAGAGTCCCAAGTTTGTGTAGGA 1320  
DB 1372 ACTATGTGAGATCAACCGGAGAAAGTACTGCGGTGAGAGTCCCAAGTTTGTGTAGGA 1431  
QY 1321 GGAACAGCAGCAAGATTACATGTCACCTTCTATCTGTATCACTGTACACGACACCGGCT 1380  
DB 1432 GGAACAGCAGCAAGATTACATGTCACCTTCTATCTGTATCACTGTATCAACGACACCGGCT 1491  
QY 1381 TCTTACTGAGTACTCTCTCTACGACTCCCAAGACCCCGCCCAAGAGATTCTCATGTGCA 1440  
DB 1492 TCTTACTGAGTACTCTCTCTACGACTCCCAAGACCCCGCCCAAGAGATTCTCATGTGCA 1551  
QY 1441 AGACTGAGCGGTGATCCGAAGAAAGTACGCTGCGGAGCGGCTGCGGAGACTCCCGGATT 1500  
DB 1552 AGACTGAGCGGTGATCCGAAGAAAGTACGCTGCGGAGCGGCTGCGGAGACTCCCGGATT 1611  
QY 1501 ATAGTATGAGGCTTACTGCGGATGCAATGCAACCCACCAAGTTCACTGTCAAAAAACAAT 1560  
DB 1612 ATAGTATGAGGCTTACTGCGGATGCAATGCAACCCACCAAGTTCACTGTCAAAAAACAAT 1671  
QY 1561 TCTGCAAGCCCTCTTCTGCGGCTGTGACAGTGTCAAGCTGTGGGAGCGGAAGTGAAG 1620  
DB 1672 TCTGCAAGCCCTCTTCTGCGGCTGTGACAGTGTCAAGCTGTGGGAGCGGAAGTGAAG 1731  
QY 1621 AGGAGGCTGACGCTCTCTGTGAGGATTCAATGTTCCAAATGGAAAGTCTCCCTC 1680  
DB 1732 AGGAGGCTGACGCTCTCTGTGAGGATTCAATGTTCCAAATGGAAAGTCTCTCCCTC 1791  
QY 1681 AGAGCCAGAAATGTATGGAAGGACAACTGTGAGATGCGTCTTGACAGAGCTTCATGTG 1740  
DB 1792 AGAGCCAGAAATGTATGGAAGGACAACTGTGAGATGCGTCTTGACAGAGCTTCATGTG 1851  
QY 1741 ACAGCGTGAATGTGTCTCTTGTGACCAAAATATACCTTACCGCTGCGCAAAATGGCTCTGTG 1800  
DB 1852 ACAGCGTGAATGTGTCTCTTGTGACCAAAATATACCTTACCGCTGCGCAAAATGGCTCTGTG 1911  
QY 1801 TGAGCAAGGCGCAACCTGAGTGTATGGAAGGACGACTGTAGGATGGCTCCGATGAGA 1860  
DB 1912 TGAGCAAGGCGCAACCTGAGTGTATGGAAGGACGACTGTAGGATGGCTCCGATGAGA 1971  
QY 1861 AAAAATGTGACTGTGGGCTGTGCAATCTTTTACAAAACAGGCTGCGGTGTGTGTGCGACA 1920  
DB 1972 AAAAATGTGACTGTGGGCTGTGCAATCTTTTACAAAACAGGCTGCGGTGTGTGTGCGACA 2031  
QY 1921 ATGCGGAGGAGGCGAGTGTGCTGTGACGTGTGACGCTTCAAGGCTTGGGCGCAAGGCGCACT 1980

DB 2032 ATGCGGAGGAGGCGAGTGTGCTCCCTGGCAGGTGAGCTCCACGCGCTGGGCGAAGGCGCACT 2091  
QY 1981 TGTGTGGGCGCTGGCTCATCTCTCTGATGCTGCTGTCTCTGTGACCTCATTTGCTTTACAG 2040  
DB 2092 TGTGTGGGCGCTGGCTCATCTCTCTGATGCTGCTGTCTCTGTGACCTCATTTGCTTTACAG 2151  
QY 2041 ATGACAAAAATTTCAAGTACTGACATCAACGATGTGAGAGGCGCTTCCGGGCTGTGTG 2100  
DB 2152 ATGACAAAAATTTCAAGTACTGACATCAACGATGTGAGAGGCGCTTCCGGGCTGTGTG 2211  
QY 2101 ACCAGAGCAAGCGAGTGTGCTTGTGGGTGAGAGGCTGAAAGCTCAACAGTATATCAACC 2160  
DB 2212 ACCAGAGCAAGCGAGTGTGCTTGTGGGTGAGAGGCTGAAAGCTCAACAGTATATCAACC 2271  
QY 2161 ACCCTTCTTCAATGATTTTCACTTGTGACTATGACATCGCTTGTGTGAGCTGGAAGT 2220  
DB 2272 ACCCTTCTTCAATGATTTTCACTTGTGACTATGACATCGCTTGTGTGAGCTGGAAGT 2331  
QY 2221 CGGTGAGTACAGACCGTGTGCGGCCCATCTGCGCTGTGATGCTACCAATGTCTTCC 2280  
DB 2332 CGGTGAGTACAGACCGTGTGCGGCCCATCTGCGCTGTGATGCTACCAATGTCTTCC 2391  
QY 2281 CTGCTGGCAAGGCGCATCTGTGATCAAGGCTGTGGGCGACAAAGAGGAGGTACCGAG 2340  
DB 2392 CTGCTGGCAAGGCGCATCTGTGATCAAGGCTGTGGGCGACAAAGAGGAGGTACCGAG 2451  
QY 2341 CGGTGATCTGAGAAAGGTGATCCGTGTATCAACGACACCTGTGAGACCTCA 2400  
DB 2452 CGGTGATCTGAGAAAGGTGATCCGTGTATCAACGACACCTGTGAGACCTCA 2511  
QY 2401 TGGCGCAGAGATCAACCCCAAGATGTGTGTGGTCTCTCATGTGGGCTGTGTGAGT 2460  
DB 2512 TGGCGCAGAGATCAACCCCAAGATGTGTGTGGTCTCTCATGTGGGCTGTGTGAGT 2571  
QY 2461 CTGCTCAAGGTGATCTGTGTGCGCTTGTCAAGGCGGAGAAAGATGGGCGAATGTTCC 2520  
DB 2572 CTGCTCAAGGTGATCTGTGTGCGCTTGTCAAGGCGGAGAAAGATGGGCGAATGTTCC 2631  
QY 2521 AGGCTGTGTGTGAGTGTGGGCTGAAAGCTGTGCTCAAGAGAACCAAGCGGCTGTACA 2580  
DB 2632 AGGCTGTGTGTGAGTGTGGGCTGAAAGCTGTGCTCAAGAGAACCAAGCGGCTGTACA 2691  
QY 2581 CAAGGCTCCCTGTGTGTGGGACTGTGATCAAGAGACACACTGGGGTATATGAGCAATGAC 2640  
DB 2692 CAAGGCTCCCTGTGTGTGGGACTGTGATCAAGAGACACACTGGGGTATATGAGCAATGAC 2751  
QY 2641 AGACAGCCGACCAACCAACCAAGGATCCCGACATGACACCTGTGATACAGAGAG 2700  
DB 2752 AGACAGCCGACCAACCAACCAAGGATCCCGACATGACACCTGTGATACAGAGAG 2811  
QY 2701 GAAACCTGACGACATTTATGTGTGTGCTCCCCCCCCCAACCAACCAAGCTGTGAAT 2760  
DB 2812 GAAACCTGACGACATTTATGTGTGTGCTCCCCCCCCCAACCAACCAAGCTGTGAAT 2871  
QY 2761 GCATCTTTAGACTCAGAGTCTTCAAAATGTGGGACCCCTCAAGAGTTGAGAGAGAC 2820  
DB 2872 GCATCTTTAGACTCAGAGTCTTCAAAATGTGGGACCCCTCAAGAGTTGAGAGAGAC 2931  
QY 2821 TTGGGTGTAGCGGCGCCAGCTGTGGGCGAAGGTTTGAATGGACCTTCCCTCTAGCC 2880  
DB 2932 TTGGGTGTAGCGGCGCCAGCTGTGGGCGAAGGTTTGAATGGACCTTCCCTCTAGCC 2991  
QY 2881 CTGAGCTGGGTGAAGATGTGTCTCCCGAGAGACTGTTCCACTGTATGAGTCTCC 2940  
DB 2992 CTGAGCTGGGTGAAGATGTGTCTCCCGAGAGACTGTTCCACTGTATGAGTCTCC 3051  
QY 2941 GGGAGCCCTATGTGGAGGAGGCGCTCAGGGTCACTTTTTCAGAAAGCGCACGCTTGA 3000  
DB 3052 GGGAGCCCTATGTGGAGGAGGCGCTCAGGGTCACTTTTTCAGAAAGCGCACGCTTGA 3111  
QY 3001 ACCCCGAAAAAGGTGTACTTAAAGCTGAATTTTGTCTGTGCGAGGGGTGGTAT 3060

Db	3112	ACCOCAGAAAAGAGTGTACTTAAGCTGAATTTGTTTCTGTGGCAGAGGTGGGTAT	3177
QY	3061	TTGAGAGTAAACATTTATTTCTTTTAAAAA	3094
Db	3172	TTGAGACTAAACATTTATTTCTTTTAGGTAA	3205
RESULT 3			
AY419858			
LOCUS			
DEFINITION	Mus musculus ST14 gene, VIRTUAL TRANSCRIPT, partial sequence,		
ACCESSION	AY419858		
VERSION	AY419858.1		
KEYWORDS	SSS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2348) Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smnsky,J.J., Adams,M.D. and Cargill,M.		
TITLE	Interfering neutral evolution from human-chimp-mouse orthologous gene trios		
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBMED	14671302		
REFERENCE	2 (bases 1 to 2348)		
AUTHORS	Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smnsky,J.J., Adams,M.D. and Cargill,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA		
COMMENT	This sequence as made by sequencing genomic exons and ordering them based on alignment.		
FEATURES	location/Qualifiers		
source	1..2348		
gene	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10090"		
	<1..>2348		
	/gene="ST14"		
	/locus_tag="HMCW7030"		
ORIGIN			
	Query Match 57.6%; Score 1790; DB 29; Length 2348;		
	Best Local Similarity 78.0%; Pred.No.0;		
	Matches 1939; Conservative 0; Mismatched 409; Indels 139; Gaps 1		
QY	144	AACATGAATGGCTTGTGAGAGGGGTGGAGTTCCTGCTGCGAACAATGCCAAGAAAGTG	203
Db	1	AACATGAATGGCTTGTGAGAGGGGTGGAGTTCCTGCTGCGAACAATGCCAAGAAAGTG	60
QY	204	GAGAAAGCGAGGCCCGCAGCGCTGGATGTGCTGTGGAGAGTGTCACTTCTCTTG	263
Db	61	GAGAAAGCGAGGCCCGCAGCGCTGGATGTGCTGTGGAGAGTGTCACTTCTCTTG	120
QY	264	CTTCTCCCTCAATGGCTGGCTTGGTGTGGCACTTCATTATCGGAATGGGGGTCAA	323
Db	121	CTTCTCCCTCAATGGCTGGCTTGGTGTGGCACTTCATTATCGGAATGGGGGTCAA	180
QY	324	AAGGTCTCAATGGCCATCTGAGATCACAATAGATCTTTCTGATCGATAGAAAC	383
Db	181	AAGGTCTCAATGGCCATCTGAGATCACAATAGATCTTTCTGATCGATAGAAAC	240
QY	384	TCCACCTCCACAGAGTTTATCAAGCTGGCCAGCCAGGTGAAGAGCGCTGAAGCTGCTG	443
Db	241	TCCACCTCCACAGAGTTTATCAAGCTGGCCAGCCAGGTGAAGAGCGCTGAAGCTGCTG	300
QY	444	TACATGAAGTCCCTGCTGCTGGGTCCCTACCAACAAGAGTGGGCTTAAGCTTCAGT	503

Db	301	TAAATGAAGTCCCTGTCCTGGGTCCCTTCCACAGAAATCGGCTGTAACTGCTTCAG-	359
Qy	504	GAGGGCAGTGCATGCGCTTAATACTGATCAGATTGAGATCCCCCACAACCTGGCAGAA	563
Db	360	-----	359
Qy	564	GAGGTGATGGCCCATGCGTGTGGAGCGAGATTGTAACTTGCAACCCGAGACAAGGACA	623
Db	360	-----	359
Qy	624	CTGAATCTTCGTGCTAACTATCTGTGTGTGCTTCGCCATGTGACCCCAAGATGCTCAG	683
Db	360	-----NN	401
Qy	684	AGGACTCAGACAACAAGCTGCAAGTTTGGCTGCAATGCCATGTGTGACACAGTGAACGCG	743
Db	402	NN	461
Qy	744	TTCACTACCCCTGGCTTCCCAACAAGCTCAACCGGCGATGCCCCGCTGCCAGTGGCTC	803
Db	462	NN	521
Qy	804	CTGGCGGGGAGACGCCGACTGTGTCTGAGGCTCACTCCGAAGCTTGTATGTGCTGCC	863
Db	522	NN	581
Qy	864	TGTATATAGCATGGCAGTGAAGCTGTGACCGGTATGATGAGCTGAGCCCATGGAACCC	923
Db	582	TGTATATAGCATGGCAGTGAAGCTGTGACCGGTATGATGAGCTGAGCCCATGGAACCC	641
Qy	924	CACGCTGTGTGCGGCTGTGTGGACCTTCTCAACCTCCTACAACTGACTTTCCTCTCC	983
Db	642	CACGCTGTGTGCGGCTGTGTGGACCTTCTCAACCTCCTACAACTGACTTTCCTCTCC	701
Qy	984	TCCCAAGAACGCTTCTCTTGTGACGCTGATTAACAATCTGACCGGAGCATCTGTGCTTT	1043
Db	702	TCCCAAGAACGCTTCTCTTGTGACGCTGATTAACAATCTGACCGGAGCATCTGTGCTTT	761
Qy	1044	GAGGCACTTCTCTTCCAGCTGCCCCAAGATGAGCAGCTGTGGCGGCTTTTGAAGTGAACCC	1103
Db	762	GAGGCACTTCTCTTCCAGCTGCCCCAAGATGAGCAGCTGTGGCGGCTTTTGAAGTGAACCC	821
Qy	1104	CAAGGACATTTTAAAGACGCTCTATATCCAGGCCACTAACCCGCCAATCACTCACTGACA	1163
Db	822	CAAGGACATTTTAAAGACGCTCTATATCCAGGCCACTAACCCGCCAATCACTCACTGACA	881
Qy	1164	TGGAATATCAAGGTGCCCAACAACCGGAAGGTGAGCTTCAAACTCTTACTGCTG	1223
Db	882	TGGAATATCAAGGTGCCCAACAACCGGAAGGTGAGCTTCAAACTCTTACTGCTG	941
Qy	1224	GTGACCCCAACGTACAGTGGGCTCTGTGCAACAAGACTATGTGGAGATCAACGGGAG	1283
Db	942	NN	1001
Qy	1284	AAGTACTGCGGGGAGAGGTCCAGTGTGTGTGAGCAGCAACAGACAAGATTACAGTC	1343
Db	1002	NNGTACTGCGGGGAGAGGTCCAGTGTGTGTGAGCAGCAACAGACAAGATTACAGTC	1061
Qy	1344	CACCTTCATCTGATCACTGTGTAACGAGACAACGAGGTTCTAGTGTGATCTTCTCTAC	1403
Db	1062	CACCTTCATCTGATCACTGTGTAACGAGACAACGAGGTTCTAGTGTGATCTTCTCTAC	1121
Qy	1404	GACTCCACGACCCGCTGCCAGGAGTGTCACTGTGCAAGACTGACGCTGATCCGAAG	1463
Db	1122	GACTCCACGACCCGCTGCCAGGAGTGTCACTGTGCAAGACTGACGCTGATCCGAAG	1181
Qy	1464	GAACTGCGCTGGACGGCTGGGCAACTGCCGGAATTAGTATGACGCTTACTGCCGA	1523
Db	1182	NN	1241
Qy	1524	TGCATGCCACCCACAGTTCACTGCAAAAACAGTTTGTGAAGCCCTCTTCTGGGTC	1583



```

Db      1242  TGCATGCCAACCCACAGTTCACGTCGCAAAAACCAAGTTCTGCAAGCCCTCTTCTGSGGTC 1301
Qy      1584  TGTGACAGTGTCAACGACTGTGGGAGCGAAGTACAGAGAGGCGCTGACGTCTCTGCT 1643
Db      1302  TGTGACAGTGTCAACGACTGTGGGAGCGAAGTACAGAGAGGCGCTGACGTCTCTGCT 1361
Qy      1644  GGGAGTTTCAAGTGTTCCTCAATGGGAAGTGTCTCCCTCAGAGCCAGAGTGTATATGGAG 1703
Db      1362  GGGAGTTTCAAGTGTTCCTCAATGGGAAGTGTCTCCCTCAGAGCCAGAGTGTATATGGAG 1421
Qy      1704  GACAACCTGTGAGATGGGTCTGACGAGGCTTCAATGTGACAGCGTGAATGTCTCTTGC 1763
Db      1422  GACAACCTGTGAGATGGGTCTGACGAGGCTTCAATGTGACAGCGTGAATGTCTCTTGC 1481
Qy      1764  ACCAATATATACCTACCGCTGCCAAATAGCCTCTGTCTGACGAGGCGCAACCTGTAGTGT 1823
Db      1482  ACCAATATATACCTACCGCTGCCAAATAGCCTCTGTCTGACGAGGCGCAACCTGTAGTGT 1541
Qy      1824  GATGGAGAGACGGAAGTGTAGCGGATGGCTCCGATGAGAGAAAACGTGACTGTGGGCTGCCA 1883
Db      1542  GATGGAGAGACGGAAGTGTAGCGGATGGCTCCGATGAGAGAAAACGTGACTGTGGGCTGCCA 1601
Qy      1884  TCCCTTACCAAAACAGGCTGCGGTGTGGTGTGACAGAAATGCGAGCGAGGCGAGTGGCC 1943
Db      1602  TCCCTTACCAAAACAGGCTGCGGTGTGGTGTGACAGAAATGCGAGCGAGGCGAGTGGCC 1661
Qy      1944  TGGGAGGTGAGCTCCAGCGCCCTGGGCGCGAGGCGCACTTGTGTGGGCTCTGCTCATCTCT 2003
Db      1662  TGGGAGGTGAGCTCCAGCGCCCTGGGCGCGAGGCGCACTTGTGTGGGCTCTGCTCATCTCT 1721
Qy      2004  CCTGACTGAGCTGTCTCTGACGCTCATGTCTTTCAGATGAGCAAAAATTTCAAGTACTGA 2063
Db      1722  CCTGACTGAGCTGTCTCTGACGCTCATGTCTTTCAGATGAGCAAAAATTTCAAGTACTGA 1781
Qy      2064  GACTACACGATGTGACGCGCTTCTCTGAGTGTCTGTCGACCAAGAGCGAGTGTCTCT 2123
Db      1782  GACTACACGATGTGACGCGCTTCTCTGAGTGTCTGTCGACCAAGAGCGAGTGTCTCT 1841
Qy      2124  GGGGGTGCAGGAGCTGAAGCTCAAGATCATATCATACCCACCTTCTTCAATATTTTACC 2183
Db      1842  GGGGGTGCAGGAGCTGAAGCTCAAGATCATATCATACCCACCTTCTTCAATATTTTACC 1901
Qy      2184  TTGCACTATGACATGCGCTTGTCTGAGAGCTGAGAAAGTGGTGGATGACGACCGCTCGG 2243
Db      1902  TTGCACTATGACATGCGCTTGTCTGAGAGCTGAGAAAGTGGTGGATGACGACCGCTCGG 1961
Qy      2244  CGCCCATCTGCTGCTGCTGATGCTTACCAATGTCTTCCCTGCTGGAGGCGCATCTGGGTC 2303
Db      1962  CGCCCATCTGCTGCTGCTGATGCTTACCAATGTCTTCCCTGCTGGAGGCGCATCTGGGTC 2021
Qy      2304  ACAGGCTGGGGGCAACAAAAAGAGAGGTACCGAGGCGCTGATCTCTGACAGAGGCTGAG 2363
Db      2022  ACAGGCTGGGGGCAACAAAAAGAGAGGTACCGAGGCGCTGATCTCTGACAGAGGCTGAG 2081
Qy      2364  ATCCGTGTATCAACCAACCAACCTGTGAGAGCTCATAGCGGCAACATATACCCACAGA 2423
Db      2082  ATCCGTGTATCAACCAACCAACCTGTGAGAGCTCATAGCGGCAACATATACCCACAGA 2141
Qy      2424  ATGATGTGTGTGGTTTCTCAAGTGGGGGTGTGAGCTCTGCAAGGCTGATCTTGTGGTGC 2483
Db      2142  ATGATGTGTGTGGTTTCTCAAGTGGGGGTGTGAGCTCTGCAAGGCTGATCTTGTGTGGTGC 2201
Qy      2484  CCGCTTGTCAAGCGCGGAGAAAGATGTGGATGTTCAGGCTGTGTGTGTGAGCTGGGGT 2543
Db      2202  CCGCTTGTCAAGCGCGGAGAAAGATGTGGATGTTCAGGCTGTGTGTGTGAGCTGGGGT 2261
Qy      2544  GAAAGCTGTGCTGTGAGAGAAACAAGCAGGCGGTGTACAACAAGCTCCCTGTAGTTCGGAG 2603
Db      2262  GAAAGCTGTGCTGTGAGAGAAACAAGCAGGCGGTGTACAACAAGCTCCCTGTAGTTCGGAG 2321
Qy      2604  TGAATCAAAAGACACTGGGGGTATAG 2630
Db      2322  TGAATCAAAAGACACTGGGGGTATAG 2348

```

```

RESULT 4
AY419856
LOCUS
DEFINITION
Homo sapiens ST14 gene, VIRTUAL TRANSCRIPT, partial sequence.
ACCESSION
AY419856
VERSION
AY419856.1 GI:39775813
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2487)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source
1..2487
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2487
/gene="ST14"
/locus_tag="HCM7030"
ORIGIN
Query Match 50.2%; Score 1558.6; DB 29; Length 2487;
Best Local Similarity 71.3%; Pred. No. 2,6e-295;
Matches 1774; Conservative 0; Mismatches 713; Indels 0; Gaps 0;
Qy 144 AACATGATGCTTTGAGAGGAGTGTGAGATTCTGCTGCGAACATGCGAAGAAAGTG 203
Db 1 AAGGTGAATGGCTTGGAGAGGAGGCGTGGAGTTCTGCGACATCAACAGTCAAGAAAGTG 60
Qy 204 GAGAGCGAGGCGCCAGGCGCGTGGGTGGTGTGCTGTGTGAGAGTGTCTTACGTTCTCTTGG 263
Db 61 GAAGAGCATGCGCGGGGCGCGTGGGTGTGCTGTGAGGCGGTGTGATGGGCTCTCTTGG 120
Qy 264 CTCTCCCTCATGAGGCTGTGCTGTGTGAGCACTTCAATTAATGGAATGTGGGGTTCA 323
Db 121 GTCTTGTGCGGGAGATGCGCTTCTGTGTGAGCAATTTGCAATACCGAGAGTGTGCTCAG 180
Qy 324 AAGGCTTCAATGGCCATCTGAGATCACAATATGATCTTTCTGATGCGTATGAGAAC 383
Db 181 AAGGCTTCAATGGCCATCTGAGATCACAATATGATCTTTCTGATGCGTATGAGAAC 240
Qy 384 TCCAGCTTCACAGATTATACAGCTGGGCGAGCGAGTGAAGAGGCGGCTGAAGTGTG 443
Db 241 TCCAGCTTCACAGATTATACAGCTGGGCGAGCGAGTGAAGAGGCGGCTGAAGTGTG 300
Qy 444 TACAATGAAGTCCCTGTCTGTGGTCCCTACCAACAAGAGTGGCTGTACTGCTTCACT 503
Db 301 TACAGCGGAGTCCCATTCCTGTGGGCGCTTACCAACAAGAGTGGCTGTGACGCGCTTCACT 360
Qy 504 GAGGCGAGTGTCAATGCGCTTACTACTGTGTCAAGATTACAGATCCCGCCACACTGGCAGAA 563

```

Db 361 GAGGCGAGCGTCATGGCCCTACTACTGTCGTGAGTTCAAGCATCCCGAGAGCACTGGTGAAG 420  
 Oy 564 GAGGTTGATCGGCGCATGGCTGTGAGCGAGTTGTAACATTGCCACCCCGAGCAAGGGA 623  
 Db 421 GAGGCGGAGCGCGTCATGGCCGAGGAGCGGTAGTATGTCGCCCGCGGCGCGCTCC 480  
 Oy 624 CTGAAATCTTGTGTGTAACATCTGTGTGTGGCTTCCCATTAAGCCCGAAATGCTGCAAG 683  
 Db 481 CTGAAGTCTTGTGTGTAACCTGAGTGTGGCTTCCCGAGGATCTCAAAACAGTACAG 540  
 Oy 684 AGGACTCAGAGCAACAGTGCAGTTTGGCTGTCATGGCCATGTGTGAGCAGTGAACAGC 743  
 Db 541 AGGACCCAGAGCAACAGTGCAGCTTGGCTGTCAGGCCGCGGTGTGAGCTGAAGCCG 600  
 Oy 744 TTCTACTACCCCTGGCTTCCCGAAAGTCCCTACCCGCGCATGTCCCGCTGCGAGTGGGTC 803  
 Db 601 TTTCACACGCGCGGCTTCCCTGACAGCCCTACCCCGCTCATGCCCTGCCAGTNNNNN 660  
 Oy 804 CTGCGGGGGAGCGCGCACTGTGTGAGCTGACCTCCGAAGCTTGTATGTGCTGCC 863  
 Db 661 NNN 720  
 Oy 864 TGTGATGAGCATGGCAGTGAAGCTGTACCGGTATGATAGCTGAGCCCATGAAACC 923  
 Db 721 NNN 780  
 Oy 924 CAGCTGTGTGCGCTGTGTGCACTTTCACCTCTCTACAACTGACTTTCCTCTCC 983  
 Db 781 NNN 840  
 Oy 984 TCCCAAGACGTCTCTCTGTGCAAGCTGATACCAATACTGACCGGAGCACTCTGCTT 1043  
 Db 841 TCCCAAGACGTCTCTGTGCAAGCTGATACCAATACTGACCGGAGCACTCTGCTT 900  
 Oy 1044 GAGGCACTTCTCTGCAAGCTGCAAGATGAGCAGCTGTGTGCGCTTTTGTGATGACCC 1103  
 Db 901 GAGGCACTTCTCTGCAAGCTGCAAGATGAGCAGCTGTGTGCGCTTTTGTGATGACCC 960  
 Oy 1104 CAAAGGACATTTAGCAGCCCTTACTATCAAGGCCACTACCCGCGCAACTCAACTGAC 1163  
 Db 961 CAGGGGACATTTAGCAGCCCTTACTATCAAGGCCACTACCCGCGCAACTCAACTGAC 1020  
 Oy 1164 TGAATATCAAGTGCCCAACACCGGAAGTGTGCGCTTCAAACTCTTATCTG 1223  
 Db 1021 TGAATATCAAGTGCCCAACACCGGAAGTGTGCGCTTCAAACTCTTATCTG 1080  
 Oy 1224 GTGAGCCCAAGTACAGTGTGCTCTGTGACCAAGACTATGTGAGATCAACGCGGAG 1283  
 Db 1081 NNN 1140  
 Oy 1284 AAGTACTCGGTGAGAGTCCCACTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTC 1343  
 Db 1141 NNAATATCGGTGAGAGTCCCACTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTC 1200  
 Oy 1344 CACTTCACTTGTGATCACTGTGTACAGGACACCGGCTTCTAGCTGATCTCTCTAC 1403  
 Db 1201 CGCTTCACTTGTGATCACTGTGTACAGGACACCGGCTTCTAGCTGATCTCTCTAC 1260  
 Oy 1404 GACTTCAACGACCGGTGCCAGGAGTTCATGTGCAAGTGTGACGCTGATCCGAAG 1463  
 Db 1261 GACTTCAACGACCGGTGCCAGGAGTTCATGTGCAAGTGTGACGCTGATCCGAAG 1320  
 Oy 1464 GAACTGCGCTGCGAGCGCTGGGAGACTGCGCGGATTATAGTATGAGCGTTACTGCCGA 1523  
 Db 1321 NNN 1380  
 Oy 1524 TGCAATGCAACCAAGTTCACGTGCAAAACAGATTCTGCAAGCCCTTCTGTGGTC 1583  
 Db 1381 TGCAATGCAACCAAGTTCACGTGCAAAACAGATTCTGCAAGCCCTTCTGTGGTC 1440  
 Oy 1584 TGTGACAGTGTCAACGACTGTGTGGGACGAAAGTGAAGGAGGCTGTGAGCTGTCTGCT 1643  
 Db 1441 TGTGACAGTGTCAACGACTGTGTGGGACGAAAGTGAAGGAGGCTGTGAGCTGTCTGCT 1500

Oy 1644 GGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAAGTAAATGGGAAG 1703  
 Db 1501 CAGACTTTCAGGTTTCCAAATGGGAAGTGTCTCTGAAAAGCAGAGTGAATGGGAAG 1560  
 Oy 1704 GAACAATGTGAGATGGGTCTGACAGAGCTTATGTGACAGGTGAATGTCTCTTTC 1763  
 Db 1561 GACGACTGTGGGGAAGGAGTCCAGAGGCTCTCTGCCCCAAGGTGAACGTGTCTGT 1620  
 Oy 1764 ACCAAATATACCTACCGCTGCCAAATAGGCTCTGTGTGAGCAAGGGCAACCTGAGT 1823  
 Db 1621 ACCAAATATACCTACCGCTGCCAAATAGGCTCTGTGTGAGCAAGGGCAACCTGAGT 1680  
 Oy 1824 GATGGGAAGCAGACTGTAGCAGTGGCTCCATGAGAAAACTGTACTGTGTGGCTGCGA 1883  
 Db 1681 GACGGGAAGAGAGACTGTAGCAGTGGCTCCATGAGAAAACTGTACTGTGTGGCTGCGA 1740  
 Oy 1884 TCCCTTACCAAAAGGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1943  
 Db 1741 TCAATCAGAGACAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800  
 Oy 1944 TGGCAGGTGAGGCTTCAGCGCCCTGGGCAAGGCACTTGTGTGGGCTTGTCTATCTCT 2003  
 Db 1801 TGGCAGGTGAGGCTTCAGCGCCCTGGGCAAGGCACTTGTGTGGGCTTGTCTATCTCT 1860  
 Oy 2004 CTTGACTGTGCTGTCTGTGCAAGCTCAATGCTTTCAAGATGACAAAAATTCAAGTACTCA 2063  
 Db 1861 CTTGACTGTGCTGTCTGTGCAAGCTCAATGCTTTCAAGATGACAAAAATTCAAGTACTCA 1920  
 Oy 2064 GACTACAGATGTGTGACAGGCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2123  
 Db 1921 GACTACAGATGTGTGACAGGCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
 Oy 2124 GGGGTGCAAGAGTGTGAAGCTCAAGCATGATCAACCACTTCTCTCAATGATTTTCAAC 2183  
 Db 1981 GGGGTGCAAGAGTGTGAAGCTCAAGCATGATCAACCACTTCTCTCAATGATTTTCAAC 2040  
 Oy 2184 TTTGACTATGACATGCTGCTTGTGTGAGCTGTGAGAGTGTGTGTGTGTGTGTGTGTGT 2243  
 Db 2041 TTTGACTATGACATGCTGCTTGTGTGAGCTGTGAGAGTGTGTGTGTGTGTGTGTGTGT 2100  
 Oy 2244 GCGCCCATCTGCTGCTGT 2303  
 Db 2101 GCGCCCATCTGCTGCTGT 2160  
 Oy 2304 ACAGGCTGGGGGACACAAAGAGGAGTACCGGAGCTGTATCTGTGAAAGGCTGAG 2363  
 Db 2161 ACAGGCTGGGGGACACAAAGAGGAGTACCGGAGCTGTATCTGTGAAAGGCTGAG 2220  
 Oy 2364 ATCGGTGTATCAACAGACCACTGTGAGACTCATGTCCGACAGCATCAACCCACGA 2423  
 Db 2221 ATCGGTGTATCAACAGACCACTGTGAGACTCATGTCCGACAGCATCAACCCACGA 2280  
 Oy 2424 ATGATGT 2483  
 Db 2281 ATGATGT 2340  
 Oy 2484 CCTTGTCAAGCGGAGAAAGATGGCGAATGTTCAGAGCTGTGTGTGTGTGTGTGTGTGT 2543  
 Db 2341 CCTTGTCAAGCGGAGAAAGATGGCGAATGTTCAGAGCTGTGTGTGTGTGTGTGTGTGT 2400  
 Oy 2544 GAAAGCTGCGCTGAGAGAAACAGCCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2603  
 Db 2401 GAAAGCTGCGCTGAGAGAAACAGCCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2460  
 Oy 2604 TGTATCAAAAGGACACTGGGGTATAG 2630  
 Db 2461 TGTATCAAAAGGACACTGGGGTATAG 2487

RESULT 5  
 AY419857  
 LOCUS AY419857 2487 bp DNA linear GSS 17-DEC-2003





[illegible]

RESULT 6	AK085410	LOCUS	DEFINITION	ACCESSION	VERSION
	AK085410	2689 bp	mRNA	linear	HNC 20-SEP-2003
			Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D610022P10 product:suppression of tumorigenicity 14 (colon carcinoma), full insert sequence.		
	AK085410				
	AK085410.1	GI:26102691			

**KEYWORDS**  
Mus musculus (house mouse)  
**SOURCE ORGANISM**

**REFERENCE**  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PMID 10349636

**REFERENCE**  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PMID 11042159

**REFERENCE**  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yoshida, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matsumi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

**JOURNAL MEDLINE**  
20530913  
11076861

**REFERENCE**  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
NATURE 409, 665-690 (2001)

**REFERENCE**  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
NATURE 420, 563-573 (2002)

**JOURNAL REFERENCE**  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanasaka, T., Hara, A., Hashinume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kohri, H., Kawai, D., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiiki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takekoshi, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

**TITLE**  
DIRECT SUBMISSION  
SUBMITTED (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Sukeni-cho, Tsurumi-ku, Yokochama Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

**COMMENT**  
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers

**FEATURES**

```

source
1. .2689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:D630022P10"
/db_xref="MGI:2422473"
/db_xref="taxon:10090"
/clone="D630022P10"
/tissue_type="Kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1. .2689
/note="suppression of tumorigenicity 14 (colon carcinoma) (MGI:MGI:1338881, GB|NM_011176, evidence: BLASTN, 100%, match=720)"

ORIGIN

Query Match      30.8%; Score 956.4; DB 11; Length 2689;
Best Local Similarity 76.7%; Pred. No. 3.5e-177;
Matches 1415; Conservative 0; Mismatches 1; Indels 428; Gaps 3;

1  CATGTAGACGGCTGCGCCGAGAGGACCAAGCGCTGTGAGACCGGCGATCGGACCCCAAA 60
110 CATGTAGACGGCTGCGCCGAGAGGACCAAGCGCTGTGAGACCGGCGATCGGACCCCAAA 169
61  CCATGGGTGACATGGGCGCGGAGGCGGAGGCGGCTCTGAGACCTTGGCGCGGAGAC 120
170 CCATGGGTGACATGGGCGCGGAGGCGGAGGCGGCTCTGAGACCTTGGCGCGGAGAC 229
121 TCAAGTACAACTCCGCGCTAGAGACATGATGGCTTTGAGAGGCGTGTGAGCTTCTGC 180
230 TCAAGTACAACTCCGCGCTAGAGACATGATGGCTTTGAGAGGCGTGTGAGCTTCTGC 289
181 CTGCGAACAATGCCAAGAAAGTGAAGACGAGCGCCGAGCGCTGGGTGTGTCTGTG 240
290 CTGCGAACAATGCCAAGAAAGTGAAGACGAGCGCCGAGCGCTGGGTGTGTCTGTG 349
241 CAGTCTGTGACGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
350 CAGTCTGTGACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 409
301 ATTATCGGAATGCGGGGTTCAAAAAGTTCATATGCGCATGTGAGATCAAAATGAGA 360
410 ATTATCGGAATGCGGGGTTCAAAAAGTTCATATGCGCATGTGAGATCAAAATGAGA 469
361 TCTTCTGTGATGCGATGAGAACTCCACTCCACAGATTTATCAG-CCTGGCCAGCCAG 419
470 TCTTCTGTGATGCGATGAGAACTCCACTCCACAGATTTATCAGCCCTGGCCAGCCAG 529
420 GTGAAGAGGCGCTGAAAGCTGTGTAATGAAAGTCCCTGTCTGGGTCCCTACCAAG 479
530 GTGAAGAGGCGCTGAAAGCTGTGTAATGAAAGTCCCTGTCTGGGTCCCTACCAAG 589
480 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 539
590 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 649
540 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 599
650 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 709
600 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 659
710 AAGTGGGTGTAACTGCTTCAATGAGGCGATGCACTGCTTAACTTATGTCAGAGTTC 769
660 -----
770 GGTGAGTGGGGGCAAAACAGGGTGGGAATGTGTGGGCGAGCCTGTAAAGGGGCGGGCT 829
660 -----
830 GGAAGAGCACACTACAGGCCCAAGGTGTCTGTGAGAGACTGTGCTGTGAGTGGCACT 889

```

```

QY 660 ----- 659
Db 890 CCTTTTGGTAATAGGAGGCAAGATTAAGAATATGTCGTGGTAATAATGTTGCTT 949
QY 660 -----CCATTAACCCAGAAATGCTGACAGAGACTTACAGACA----- 696
Db 950 TCATGTTCTCAACCCATTACCCAGAAATGCTGACAGAGACTTACAGAGATATCAT 1009
QY 697 ----- 696
Db 1010 GGTGACCTGTGGTGAAGAAAAAGTGTCTTGCCCTTAAGCCATGAGGCTCTAAAGT 1069
QY 697 ----- 696
Db 1070 CAGATGTACTGATGACAAAGATCTCTTTTCCCTTCCGCTCCAAAGCTCTTTC 1129
QY 697 ----- 696
Db 1130 TTGTGTACTGATGATTCGAGTGTGATGAGACGTAATGGGGGTAGAGCTAGGCGGGG 1189
QY 697 -----ACAGTGCAGTTTTC 712
Db 1190 CGGGGCGCTGACGGGTTACAGCTTCTTCCATCTCTCCACAGACAGCTGAGTTTTC 1249
QY 713 CCTGATGCCATGTGTGACAGATGACAGCTTCACTACCTGCTTCCCAACAGTTC 772
Db 1250 CCTGATGCCATGTGTGACAGATGACAGCTTCACTACCTGCTTCCCAACAGTTC 1309
QY 773 CTACCCGCGCATGTCCTGCTGCACTGAGTCTGCGGGGAGCGCCGACTGTGTGAG 832
Db 1310 CTACCCGCGCATGTCCTGCTGCACTGAGTCTGCGGGGAGCGCCGACTGTGTGAG 1369
QY 833 CCTACCTTCCGAACCTTGAATGTGCTTCCCTGATGAGATGACATGACAGTGTGAC 892
Db 1370 CCTACCTTCCGAACCTTGAATGTGCTTCCCTGATGAGATGACATGACAGTGTGAC 1429
QY 893 CGTGTATGATAGCTTGAAGCCCATGAGAACCCACGCTGTGTGCGGCTGTGACCTT 952
Db 1430 CGTGTATGATAGCTTGAAGCCCATGAGAACCCACGCTGTGTGCGGCTGTGACCTT 1489
QY 953 CTCACCTTCTTCAACCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1012
Db 1490 CTCACCTTCTTCAACCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1549
QY 1013 AACCAATCTGACCGGCGAATCCGCTTGAAGGCGCTTCTTCCAGCTGCGCCAAAT 1072
Db 1550 AACCAATCTGACCGGCGAATCCGCTTGAAGGCGCTTCTTCTTCTGACGCTGAT 1609
QY 1073 GAGCAGCTGTGCGGCTTTTGAAGTACACCCAAAGGACATTTAGCAGCCCTATCTTC 1132
Db 1610 GAGCAGCTGTGCGGCTTTTGAAGTACACCCAAAGGACATTTAGCAGCCCTATCTTC 1669
QY 1133 AGGCCACTACCCGCGCAATCAACTGCAATGCAATGCAATGCAAGTGTCCCAACCGGAA 1192
Db 1670 AGGCCACTACCCGCGCAATCAACTGCAATGCAATGCAAGTGTCCCAACCGGAA 1729
QY 1193 CGTGAAGTGTGCTTAACTCTTCTATCTGAGTGAAGCCCAAGCAAGTGTGCTCTTC 1252
Db 1730 CGTGAAGTGTGCTTAACTCTTCTATCTGTGAGTGAAGCCCAAGCAAGTGTGCTCTTC 1789
QY 1253 CACCAAGCATATGTGAGATCAACGAGGAGAACTAGCGGTGAGAGTCCCAAGTTGT 1312
Db 1790 CACCAAGCATATGTGAGATCAACGAGGAGAACTAGCGGTGAGAGTCCCAAGTTGT 1849
QY 1313 GGTGAGCAGCAACGACAGCAAGATTAAGTCACTTCCATTTCTGATCACTGTACAGGA 1372
Db 1850 GGTGAGCAGCAACGACAGCAAGATTAAGTCACTTCCATTTCTGATCACTGTACAGGA 1909
QY 1373 CACCGGTTCTAGCTAGTACCTCTCTTACAGCTCCCAAGAGC 1416
Db 1910 CACCGGTTCTAGCTAGTACCTCTCTTACAGCTCCCAAGAGC 1953

```

RESULT 7  
 BQ885551 944 bp mRNA linear EST 16-ANG-2002  
 LOCUS BQ885551  
 DEFINITION AGENCOURT 8748027 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6335237  
 5', mRNA sequence.  
 ACCESSION BQ885551  
 VERSION BQ885551.1 GI:22277569  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 944)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
 Ph.D. Library Preparation: Resgen, Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LHAM13797 row: f column: 06  
 High quality sequence stop: 623.  
 Location/Qualifiers  
 1..944  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6335237"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organism: oocytes; Vector: pCMV-SPORT6.1; Site: 1;  
 EcoRV; Site 2: NotI; Cloned unidirectionally. Primer:  
 Oligo dT. Average insert size 1.95 kb. Constructed by  
 Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."  
 ORIGIN  
 Query Match 27.2%; Score 843.8; DB 13; Length 944;  
 Best Local Similarity 96.3%; Pred. No. 4.8e-155;  
 Matches 885; Conservative 0; Mismatches 32; Indels 2; Gaps 2;  
 Db 1 GCGCTGAGACCGGCGATCGGACCGCAAAACCATGGGTAGCAATCGGCGCGCAAGGCC 89  
 QY 30 GCGCTGAGACCGGCGATCGGACCGCAAAACCATGGGTAGCAATCGGCGCGCAAGGCC 89  
 QY 90 GAGGGGGGCTCTCAGGACTTCGGGGGGGACTCAAGTACAACTCCGGGTAGAGAAATG 149  
 Db 61 GAGGGGGGCTCTCAGGACTTCGGGGGGGACTCAAGTACAACTCCGGGTAGAGAAATG 120  
 QY 150 AATGGCTTTGAGAGGGGTGAGAGTCTCGGCTCGCAACAATGCGCAAGAAAGTGAAG 209  
 Db 121 AATGGCTTTGAGAGGGGTGAGAGTCTCGGCTCGCAACAATGCGCAAGAAAGTGAAG 180  
 QY 210 CGAGGCCCGAGGCGCTGGGTGGTCTGGTGGCACTGCTTCACTTCTCTCTCC 269  
 Db 181 CGAGGCCCGAGGCGCTGGGTGGTCTGGTGGCACTGCTTCACTTCTCTCTCC 240  
 QY 270 CTATGGCTGGCTTGGTGGTGGCACTTCCATTATGGAGATGCGGGGTCAAAAAGTC 329  
 Db 241 CTATGGCTGGCTTGGTGGTGGCACTTCCATTATGGAGATGCGGGGTCAAAAAGTC 300  
 QY 330 TTCAATGGCCATCGAGATCAAAATGATCTTTCTGATGGGTATGAGACTCCACC 389  
 Db 301 TTCAATGGCCATCGAGATCAAAATGATCTTTCTGATGGGTATGAGAACTCCACC 360  
 QY 390 TCACAGAGTTTATCAGCTTGGCCAGCCAGGTGAAGAGGCGCTGAAGCTCTGTCAAT 449  
 Db 361 TCACAGAGTTTATCAGCTTGGCCAGCCAGGTGAAGAGGCGCTGAAGCTCTGTCAAT 420

QY 450 GAAATCCCTGCTCTGAGGTCCCTACCAAGAAGTGGCTGTAACTGCTTCACTAGTAGGC 509  
 Db 421 GAAATCCCTGCTCTGAGGTCCCTACCAAGAAGTGGCTGTAACTGCTTCACTAGTAGGC 480  
 QY 510 AGTGTATCGCTACTACTACTGTGTAAGATTGACGATCCCTCCACACTGGCAAGAGGTT 569  
 Db 481 AGTGTATCGCTACTACTACTGTGTAAGATTGACGATCCCTCCACACTGGCAAGAGGTT 540  
 QY 570 GATCGGGCCATGGCTGTGAGAGGAGTTGTAACATTGGCCACCCCGAGCAAGGCACTGA 629  
 Db 541 GATCGGGCCATGGCTGTGAGAGGAGTTGTAACATTGGCCACCCCGAGCAAGGCACTGA 600  
 QY 630 TCTTGTGCTAAACATCTGTGTGCTTCCCATTTGACCCCAAGATGTGCAAGAGACT 689  
 Db 601 TCTTGTGCTAAACATCTGTGTGCTTCCCATTTGACCCCAAGATGTGCAAGAGACT 660  
 QY 690 CAGACACAAGCTGCACTTTTGGCTCTGCAATGCCATGCTGACAGTACAGCTTCACT 749  
 Db 661 CAGACACAAGCTGCACTTTTGGCTCTGCAATGCCATGCTGACAGTACAGCTTCACT 720  
 QY 750 ACCCTGGCTTCCCAACAGTCCCTACCCGCGCATGCCGCTGCAAGTGGTCTGCGG 809  
 Db 721 ACCCTGGCTTCCCAACAGTCCCTACCCGCGCATGCCGCTGCAAGTGGTCTGCGG 780  
 QY 810 GGGGAGCGCGACTCTGTGTGCAAGCTTCCGAGCTTTGATGTGCTCTGTGAT 869  
 Db 781 GGGGAGCGCGACTCTGTGTGCAAGCTTCCGAGCTTTGATGTGCTCTGTGAT 840  
 QY 870 GAGCAT-GGCATGACCTGTGACCGT-GTATATAGCCAGACCCCATGGAACCCACG 927  
 Db 841 GACCATGAGGATGTCCTGCGCCCGGTGTATGTAACCTGAGCCCATGGAACCCACG 900  
 QY 928 CTGTGTGGCGGCTGTGTGG 946  
 Db 901 CCTGTGTGTGCGGGCTTG 919  
 RESULT 8  
 B1688367 884 bp mRNA linear EST 18-SEP-2001  
 LOCUS B1688367  
 DEFINITION 603316592F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5356487 5',  
 mRNA sequence.  
 ACCESSION B1688367  
 VERSION B1688367.1 GI:15650996  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 884)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LHAM1906 row: h column: 24  
 High quality sequence stop: 831.  
 Location/Qualifiers  
 1..884  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5356487"  
 /sex="female, virgin"  
 FEATURES  
 source

```

/tissue type="infiltrating ductal carcinoma"
/dev stage="5 months"
/lab host="DH10B"
/clone lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

```

```

Query Match      26.8%; Score 833.6; DB 12; Length 884;
Best Local Similarity 98.6%; Pred. No. 4.8e-153;
Matches 872; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

```

```

QY 1165 GGAATATCAAGGTGCCCAACACCGGAAGCTGAAGTGCCTTCAACTCTTATCTG 1224
DB 1 GGAATATCAAGGTGCCCAACACCGGAAGCTGAAGTGCCTTCAACTCTTATCTG 60
QY 1225 TGGACCCCAACGATGAGGTGGCTCTGACCAAGGATATGATGATCAAGGGGAGA 1284
DB 61 TGGACCCCAACGATGAGGTGGCTCTGACCAAGGATATGATGATCAAGGGGAGA 120
QY 1285 AGTACTGCGGTGAGAGGTCCCAATTGTGTGAGACCAACAGCAGATTACAGTCC 1344
DB 121 AGTACTGCGGTGAGAGGTCCCAATTGTGTGAGACCAACAGCAGATTACAGTCC 180
QY 1345 ACTTCATTCTATCTACTGTACACGGAACCGGGTTCTTACTGATGATCTCTTACG 1404
DB 181 ACTTCATTCTATCTACTGTACACGGAACCGGGTTCTTACTGATGATCTCTTACG 240
QY 1405 ACTCCAGACCCGTCGCCAGGATGTATGAGCAAGCTGACCGGTGATCCGAAG 1464
DB 241 ACTCCAGACCCGTCGCCAGGATGTATGAGCAAGCTGACCGGTGATCCGAAG 300
QY 1465 AACTGCGTGCAGACGGCTGGGAGACTGCCGATTAATGATGAGCTTACTGCGAT 1524
DB 301 AACTGCGTGCAGACGGCTGGGAGACTGCCGATTAATGATGAGCTTACTGCGAT 360
QY 1525 GGAATGCCCAACCACTGACGTTGCAAAACCAAGTTCTGCAAGCCCTCTTCTGGTCT 1584
DB 361 GGAATGCCCAACCACTGACGTTGCAAAACCAAGTTCTGCAAGCCCTCTTCTGGTCT 420
QY 1585 GTGACAGTGTCAACGCTGTGGGAGACGGAAGTGAAGAGGAGGCTGCACTGCTGCTG 1644
DB 421 GTGACAGTGTCAACGCTGTGGGAGACGGAAGTGAAGAGGAGGCTGCACTGCTGCTG 480
QY 1645 GGAGTTTCAAGTGTTCATGGAAGTGTCTCCCTCAGACGAGAACTGAATGGGAAAG 1704
DB 481 GGAGTTTCAAGTGTTCATGGAAGTGTCTCCCTCAGACGAGAACTGAATGGGAAAG 540
QY 1705 ACAACTGTGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTGA 1764
DB 541 ACAACTGTGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTGA 600
QY 1765 CCAAAATATCTACCGCTGCCAAATATGGCTCTGTCTGAG-CAAGGGGAAACCTTAAATGT 1823
DB 601 CCAAAATATCTACCGCTGCCAAATATGGCTCTGTCTGAG-CAAGGGGAAACCTTAAATGT 660
QY 1824 GATGGGAAGACGAGCTGTAGCGATGAGCTCCGATGAGAAAACTGTGACTGTGGCTGCA 1883
DB 661 GATGGGAAGACGAGCTGTAGCGATGAGCTCCGATGAGAAAACTGTGACTGTGGCTGCA 720
QY 1884 TCCCTTACCAACAGGCTGCGG-TGGTTGTGTGACGAATGTGGAAGAGGAGGAGTGGCC 1942
DB 721 TCCCTTACCAACAGGCTGCGG-TGGTTGTGTGACGAATGTGGAAGAGGAGGAGTGGCC 780
QY 1943 CTGGCAGGTGAGCTCCAGCGCTGCGGCAAGGCAACTTGTGTGGGCTGCGC-TCAATCT 2001
DB 781 CTGGCAGGTGAGCTCCAGCGCTGCGGCAAGGCAACTTGTGTGGGCTGCGC-TCAATCT 840
QY 2002 CTCCTGACTGAGTGTCTCTGACAGCTCAATGCTTTCAGAGATGAC 2045
DB 841 CTCCTGACTGAGTGTCTCTGACAGCTCAATGCTTTCAGAGATGAC 884

```

```

RESULT 9
BI690615 803 bp mRNA linear EST 18-SEP-2001
LOCUS 603314210F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5354067 5'
DEFINITION
mRNA sequence.
BI690615
BI690615.1 GI:15653244
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 803)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Contact: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LIML1900 row: d column: 04
High quality sequence stop: 755.
Location/Qualifiers
1..803
/organism="Mus musculus"
/mol type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5354067"
/sex="female, virgin"
/tissue type="infiltrating ductal carcinoma"
/dev stage="5 months"
/lab host="DH10B"
/clone lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

```

## FEATURES

source

```

High quality sequence stop: 755.
Location/Qualifiers
1..803

```

## ORIGIN

```

Query Match      25.4%; Score 788.8; DB 12; Length 803;
Best Local Similarity 99.6%; Pred. No. 3e-144;
Matches 801; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 654 GCTTCCCATTTGACCCCAAGATGCTGAGAGGACTCAGAGCAACAGCTGAGTTTGGC 713
DB 1 GCTTCCCATTTGACCCCAAGATGCTGAGAGGACTCAGAGCAACAGCTGAGTTTGGC 60
QY 714 CTGCATGCCATGTGTGACAGAGTGAACAGCTTCACTAACCCCTGCTCCCAACAGTCC 773
DB 61 CTGCATGCCATGTGTGACAGAGTGAACAGCTTCACTAACCCCTGCTCCCAACAGTCC 120
QY 774 TACCCGGGCAATGCCCGCTGCCAGTGGTCTGCGGGGGAGACGCCGACTCTGTGTGAGC 833
DB 121 TACCCGGGCAATGCCCGCTGCCAGTGGTCTGCGGGGGAGACGCCGACTCTGTGTGAGC 180
QY 834 CTCACCTTCGGAAGCTTGAATGTGCTCCCTGATATGAGCATGTGAGAGTACCTGTGACC 893
DB 181 CTCACCTTCGGAAGCTTGAATGTGCTCCCTGATATGAGCATGTGAGAGTACCTGTGACC 240
QY 894 GTGTATGATAGCTGAGGCCCATGGAACCCCAACAGCTGTGTGTGGCTGTGTGACCTTC 953
DB 241 GTGTATGATAGCTGAGGCCCATGGAACCCCAACAGCTGTGTGTGGCTGTGTGACCTTC 300
QY 954 TCACCCCTCTACACCTGACTTTCTCTCTCCAGAAAGTCTTCTTGTGACGCTGATA 1013

```





Db 956 CAGGGGCCCCGTGGAGGGGAGCACTCCCGGCGCTGG 990

RESULT 11  
LOCUS BG919617  
DEFINITION 762 bp mRNA linear EST 05-JUN-2001  
602823173F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:4951962 5',  
mRNA sequence.  
ACCESSION BG919617 GI:14300093  
VERSION BG919617  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 762)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
<http://image.llnl.gov>  
Plate: LLAM10909 row: a column: 19  
High quality sequence stop: 727.  
location/Qualifiers  
1..762  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4951962"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies, Investigator  
providing samples: Jeffrey Green, M.D., NIH"

ORIGIN

Query Match 24.5%; Score 760.4; DB 12; Length 762;  
Best Local Similarity 99.9%; Pred. No. 1.1e-138;  
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 682 AGAGGACTCAGAGCAACAGCTGACGTTTGGCCGTCATGCCAGTGTGCACAGTGCAC 741  
Db 1 AGAGGACTCAGAGCAACAGCTGACGTTTGGCCGTCATGCCAGTGTGCACAGTGCAC 60  
QY 742 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGCGCATGCCGCTGCCAGTGGG 801  
Db 61 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGCGCATGCCGCTGCCAGTGGG 120  
QY 802 TCCGCGGGGGGAGCCGACCTCTGTGCTGAGCTCACCCTTCGGAAGCTTTGATGTCGCTC 861  
Db 121 TCCGCGGGGGGAGCCGACCTCTGTGCTGAGCTCACCCTTCGGAAGCTTTGATGTCGCTC 180  
QY 862 CTTGTATGAGCATGGAGTGAACCTGGTACCCGTTATGATAGCTAGAGCCCATGGAAC 921  
Db 181 CTTGTATGAGCATGGAGTGAACCTGGTACCCGTTATGATAGCTAGAGCCCATGGAAC 240  
QY 922 CCAAGCTGTGTGGGCTGTGTGGACCTTCTCACCCTCTCAACACTGACTTTCTCT 981  
Db 241 CCAAGCTGTGTGGGCTGTGTGGACCTTCTCACCCTCTCAACACTGACTTTCTCT 300

QY 982 CTTCCAGAACGCTCTCTGTGACGCTGATACCAATCTGACGGGCGACATCTGGCT 1041  
Db 301 CTTCCAGAACGCTCTCTGTGACGCTGATACCAATCTGACGGGCGACATCTGGCT 360  
QY 1042 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGGCTTTTGTAGTACA 1101  
Db 361 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGGCTTTTGTAGTACA 420  
QY 1102 CCCAAGGACATTTAGCAGCCCTTACTATTCAGGCCACTACCCGCCCAATCACTGCA 1161  
Db 421 CCCAAGGACATTTAGCAGCCCTTACTATTCAGGCCACTACCCGCCCAATCACTGCA 480  
QY 1162 CATGCAATATCAAGGAGCCCAACACCGAAGTGAAGGTGGCTTCAACTTTCTATC 1221  
Db 481 CATGCAATATCAAGGAGCCCAACACCGAAGTGAAGGTGGCTTCAACTTTCTATC 540  
QY 1222 TGGTGACCCCAACGTCACAGTGGGCTCTCTGACCAAGACATATGTGAGATCAACGGGG 1281  
Db 541 TGGTGACCCCAACGTCACAGTGGGCTCTCTGACCAAGACATATGTGAGATCAACGGGG 600  
QY 1282 AGAAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGACGACACAGACCAAGATTACG 1341  
Db 601 AGAAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGACGACACAGACCAAGATTACG 660  
QY 1342 TCCACTTCATCTGTATCACTGTCACGGAACCGGGGTTCTAGTGTAGTACCTCTCT 1401  
Db 661 TCCACTTCATCTGTATCACTGTCACGGAACCGGGGTTCTAGTGTAGTACCTCTCTCT 720  
QY 1402 ACGACTCCACGACCCGTGCCAGGAGATGTTGATGTGCAGA 1443  
Db 721 ACGACTCCACGACCCGTGCCAGGAGATGTTGATGTGCAGA 762

RESULT 12  
LOCUS CB208872 874 bp mRNA linear EST 16-MAY-2003  
DEFINITION AGENCOURT\_11350695 NIH MGC 164 Mus musculus cDNA clone  
IMAGE:30241011 5', mRNA sequence.  
ACCESSION CB208872  
VERSION CB208872.2 GI:29133712  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 874)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT On Feb 4, 2003 this sequence version replaced gi:28250435.  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe and Dr. Mina  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM0308 row: g column: 04  
High quality sequence stop: 655.  
location/Qualifiers  
1..874  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30241011"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 164"  
/note="Vector: pCMV-SPORT6.1; Site\_1: EcoRV; Site\_2: NotI;  
Non-normalized full-length enriched library from pooled  
mouse embryonic limb, maxilla and mandible, day 10.5 and  
11.5 (size selected for the 0.5-1 kb fragments) Cloned

FEATURES  
source

directionally, priming method: Oligo-dT. cDNA enrichment:  
 21k bp, Average insert size 1.8k bp, Priming sequence:  
 5'-GACTAGTCTAGATCGAGGCGGCGCCCTT-3', Tissue contributed  
 by, David Rowe. Library constructed by Resgen, Invitrogen  
 Corp."

## ORIGIN

Query Match 24.4%; Score 759.4; DB 14; Length 874;  
 Best Local Similarity 98.9%; Pred. No. 1.7e-136;  
 Matches 785; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

1387 CTGAGTACCTCTCTACGACATCCAAAGACCCGTCAGGATGTTTCATGTGCAAGACTG 1446  
 1 CTGAGTACCTCTCTCTACGACATCCAAAGACCCGTCAGGATGTTTCATGTGCAAGACTG 60

1447 GACGGTGCATCCGAAAGAACTGCGCTGGAGCGCTGGGCAAGCTCCCGGATTAATAGG 1506  
 61 GACGGTGCATCCGAAAGAACTGCGCTGGAGCGCTGGGCAAGCTCCCGGATTAATAGG 120

1507 ATGAGCGTTACTGCGGATGCAATGCAACCCACCGATTCAGTGCACAAACAGATTCTGCA 1566  
 121 ATGAGCGTTACTGCGGATGCAATGCAACCCACCGATTCAGTGCACAAACAGATTCTGCA 180

1567 AGCCCTCTCTCTGCTGTGACAGTGTCAACGACTGTGGGACGGAAGTACAGAGAGG 1626  
 181 AGCCCTCTCTCTGCTGTGACAGTGTCAACGACTGTGGGACGGAAGTACAGAGAGG 240

1627 GCTGACGCTGTCTCTGCTGGAGTTTCAAGTGTTCATAGGAAAGTGTCTCCCTCAAGCC 1686  
 241 GCTGACGCTGTCTCTGCTGGAGTTTCAAGTGTTCATAGGAAAGTGTCTCCCTCAAGCC 300

1687 AGAAGTGTAATGGGAAAGCAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGG 1746  
 301 AGAAGTGTAATGGGAAAGCAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGG 360

1747 TGAATGCTGTCTCTGACCAAAATATACCTACCGCTGCCAAAATGGCTGTGTGAGCA 1806  
 361 TGAATGCTGTCTCTGACCAAAATATACCTACCGCTGCCAAAATGGCTGTGTGAGCA 420

1807 AGGCAACCTCTGAGTGTGATGGGAAAGCAAGCTGTAGCAATGCTCCGATGAGAAAACT 1866  
 421 AGGCAACCTCTGAGTGTGATGGGAAAGCAAGCTGTAGCAATGCTCCGATGAGAAAACT 480

1867 GTGACTGTGGGCTGCGATCCCTTACCAAAAGAGCTCGGCTGTGTGAGCAAGATCGG 1926  
 481 GTGACTGTGGGCTGCGATCCCTTACCAAAAGAGCTCGGCTGTGTGAGCAAGATCGG 540

1927 ACAGAGGCGAGTGGCCCTGGAGGTGAGCTCCACGCTGGGCGAGGCGCACTTGTGTG 1986  
 541 ACAGAGGCGAGTGGCCCTGGAGGTGAGCTCCACGCTGGGCGAGGCGCACTTGTGTG 600

1987 GGGCTCTGCTCATCTCTCTGACCTGGCTGTCTGACGACTATTGCTTCAGAGATGCA 2046  
 601 GGGCTCTGCTCATCTCTCTGACCTGGCTGTCTGACGACTATTGCTTCAGAGATGCA 660

2047 AAAATTTCAGTACTCAGACTACAGATGATGAGCGCTTCTGCTGCTGAGACAGA 2106  
 661 AAAATTTCAGTACTCAGACTACAGATGATGAGCGCTTCTGCTGCTGAGACAGA 719

2107 GCAAGCGCAGTGCCTCTGGGGGTGAGAGAGTGAAGCTCAAAAGTATCAT -CACTCAGCTT 2165  
 720 GCAAGCGCAGTGCCTCTGGGGGTGAGAGAGTGAAGCTCAAAAGTATCATCCGCCACCGT 779

2166 TCCTTCAATGATTT 2179  
 780 TCCTTCAATGATTT 793

RESULT 13  
 B1664081 798 bp mRNA linear EST 12-SEP-2001  
 LOCUS B1664081 60328385F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5323485 5',  
 DEFINITION mRNA sequence.

ACCESSION B1664081  
 VERSION B1664081.1 GI:15578314  
 EST.  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 798)  
 AUTHORS NIH-MGC  
 TITLE NIH-MGC http://mgc.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1M11820 row: 1 column: 22  
 High quality sequence stop: 771.  
 Location/Qualifiers

## FEATURES

## source

1..798  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /db\_xref="IMAGE:5323485"  
 /clone="female, virgin"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam6"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; site 1: SalI;  
 site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Jeffrey Green, M.D., NIH"

## ORIGIN

Query Match 24.1%; Score 747.2; DB 12; Length 798;  
 Best Local Similarity 98.6%; Pred. No. 4.3e-136;  
 Matches 785; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

16 CCGGAGGAGCAACGCTCTGAGACCGGCGATGGAATCCGCCAAACATGGGTGACATC 75  
 1 CCGGAGGAGCAACGCTCTGAGACCGGCGATGGAATCCGCCAAACATGGGTGACATC 60

76 GGGGCGCCAGGCGCGAGGGGGCTCTCAGGACTTCGGCGGAGCTCAAGTACAATCCC 135  
 61 GGGGCGCCAGGCGCGAGGGGGCTCTCAGGACTTCGGCGGAGCTCAAGTACAATCCC 120

136 GGGTGAAGACATGATGCTTTGAGAGAGGGGTGTGAGATTCCTGCTGCGAACATGCCA 195  
 121 GGGTGAAGACATGATGCTTTGAGAGAGGGGTGTGAGATTCCTGCTGCGAACATGCCA 180

196 AGAAAGTGAAGAGCGAGCGCCAGGCGCTGGGTGTGTGCTGGTGGCAAGTGTTCAGCT 255  
 181 AGAAAGTGAAGAGCGAGCGCCAGGCGCTGGGTGTGTGCTGGTGGCAAGTGTTCAGCT 240

256 TCCTCTTGTCTCTCTCATGAGCTGGCTGTGCTGGGTGTGTGCTTCATTCAGATGTGC 315  
 241 TCCTCTTGTCTCTCTCATGAGCTGGCTGTGCTGGGTGTGTGCTTCATTCAGATGTGC 300

316 GGGTTCAAAAAGTCTTCATAGGCGCATGAGAGTCAAAATGAGATCTTCTGATGTGCT 375  
 301 GGGTTCAAAAAGTCTTCATAGGCGCATGAGAGTCAAAATGAGATCTTCTGATGTGCT 360

376 ATGAGAACTCACTCCACAGAGTTTATCAGCTTGGCCAGGCCAGGTGAGAGAGGCGCTGA 435  
 361 ATGAGAACTCACTCCACAGAGTTTATCAGCTTGGCCAGGCCAGGTGAGAGAGGCGCTGA 420

436 AGCTGCTGTACATGAAATCCCTGTCTCTGGGTCTTACCAACAAGAGTGGCTGTACTG 495

Db	421	ACCTGCTGTAACAATGAAGTCCCTGTCCTCGGTCCTTACCACAAGAAGTCGGCTGTAACTG	480
Qy	496	CCCTCAGTAGAGGCGAGTGTGCATCGCCTACTACCTGGTGCAGAGTTCAGCATCCGCCCAACAC	555
Db	481	CCTTCAGTAGAGGCGAGTGTGCATCGCCTACTACCTGGTGCAGAGTTCAGCATCCGCCCAACAC	540
Qy	556	TGGCAGAGAGAGGTTGATCGCGCCATCGCTGTGGAGAGGAGTTGAACATTGACCCACCCGAG	615
Db	541	TGGCAGAGAGAGGTTGATCGCGCCATCGCTGTGGAGAGGAGTTGAACATTGACCCACCCGAG	600
Qy	616	CACGGGCGACTGAATTCCTTCGTGCTTAACATCTGTGTGGCCTTCCCATTTGACCCCGAGA	675
Db	601	CACGGGCGACTGAATTCCTTCGTGCTTAACATCTGTGTGGCCTTCCCATTTGACCCCGAGA	660
Qy	676	TGCTCAGAGAGACTCAGGACCAAGCTGACGTTTGCCCTGATG-CCCATGTGTCAGCA	734
Db	661	TGCTCAGAGAGACTCAGGACCAAGCTGACGTTTGCCCTGATG-CCCATGTGTCAGCA	720
Qy	735	GTG-ACACGCTTCACTACCCCTGCGCTTCCCAACAGTCCCTTACCCGCGCATGCCGCTG	793
Db	721	GTGAACACGCTTCACTACCCCTGCGCTTCCCAACAGTCCCTTACCCGCGCATGCCGCTG	779
Qy	794	CCAGTGGGTCTCCGCGG 809	
Db	780	GCAGTGGGTCTCCGCGG 795	
RESULT 14			
LOCUS	BY756896	746 bp	mRNA linear EST 17-DEC-2002
DEFINITION	BY756896 RIKEN full-length enriched, blastocyst Mus musculus cDNA		
ACCESSION	BY756896		
VERSION	BY756896.1	GI:27190109	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Bukacinska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 746)		
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bredt, D., Brusic, V., Chothia, C., Corradi, L. E., Cousins, S., Dalia, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Fraser, K. S., Gasteland, T., Gatholdi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kanao, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramechandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, D., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempile, C. A., Sevon, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Vearato, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, N., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carrinici, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, I., Aizawa, K., Akakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kawana, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.		
TITLE	Analysis of the mouse transcriptome based on functional annotation		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	22354683		
PUBMED	12466851		

COMMENT	FEATURES	ORIGIN
<p>Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URI: http://genome.gsc.riken.go.jp/ Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imokata, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Komuro, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Saito, R., Sakazume, N., Sano, H., Saeki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission</p> <p>Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (<a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a>) for further details. Location/Qualifiers</p>	<p>source</p> <p>1. 746 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="11C0035F06" /dev_stgae="Dlastocyst" /clone_1nb="RIKEN full-length enriched, blastocyst"</p>	<p>Query Match 23.84; Score 739.2; DB 13; Length 746; Best Local Similarity 99.38; Pred. No. 1,6e-134; Matches 741; Conservative 0; Mismatches 5; Indels 0; Gaps 0;</p>
	<p>2343 CTGATCCTGCGAAGAGGGGTGAGATCCGTGTCATCAACACGACCACTGTGAGACCTCATG 2402 1 CTGATCMTGCGAAGAGGTGAGATCCGTGTCATCAACACGACCACTGTGAGACCTCATG 60</p> <p>2403 CCGCAGACGATCAACCCACGAAATGATGTGTGTGGGTTTCTCATGTG99GGGTGTGACTCC 2462 61 CCGCAGACGATCAACCCACGAAATGATGTGTGTGGGTTTCTCATGTG99GGGTGTGACTCC 120</p> <p>2463 TGCACGAGGTATACCTGTGGTGGCCCTTGTCAAGGCGGAGAAAGATGGGGGAATGTTCCAG 2522 121 TGCACGAGGTATACCTGTGGTGGCCCTTGTCAAGGCGGAGAAAGATGGGGGAATGTTCCAG 180</p> <p>2523 GCTGTGTGTGTGAGCTGGGGGTGAGGCTGCGCTCAGAGGAAACGACGAGCGTGTACACA 2582 181 GCTGTGTGTGTGAGCTGGGGGTGAGGCTGCGCTCAGAGGAAACGACGAGCGTGTACACA 240</p> <p>2583 AGGCTCCCTGTATTTGGGACCTGGATCAAAAGACACATCGGGGTATATGACAGATGACAG 2642 241 AGGCTCCCTGTATTTGGGACCTGGATCAAAAGACACATCGGGGTATATGACAGATGACAG 300</p> <p>2643 ACAGCGACCAACAAACCCACGAGGATCCCGACATGACACCTGTGATACGAGAGAGA 2702 301 ACAGCGACCAACAAACCCACGAGGATCCCGACATGACACCTGTGATACGAGAGAGA 360</p>	

QY 2703 AACCTGACGACATTTATGCTGTGGCTCCCCCCCCCAACAACCCAGACTGGAAGCTGC 2762  
 DB 361 AACCTGACGACATTTATGCTGTGGCTCCCCCCCCCAACAACCCAGACTGGAAGCTGC 420  
 QY 2763 ATCCCTTAGACATCAGAGTCTTCTCAAGTGGAGCCCTCCTCAGAGTTGGAGAGAACTT 2822  
 DB 421 ATCCCTTAGACATCAGAGTCTTCTCAAGTGGAGCCCTCCTCAGAGTTGGAGAGAACTT 480  
 QY 2823 GCGTGTAGAGGCGCCAGAGCTGGGAGGCAAGGTTTATGAGAGCTTCCCTTACCTT 2882  
 DB 481 GCGTGTAGAGGCGCCAGAGCTGGGAGGCAAGGTTTATGAGAGCTTCCCTTACCTT 540  
 QY 2883 GAGCTGGTGAAGATATGCTGTCCCGAGAGAGCTGCTCAACTGATTTAGCTCCCG 2942  
 DB 541 GAGCTGGTGAAGATATGCTGTCCCGAGAGAGCTGCTCAACTGATTTAGCTCCCG 600  
 QY 2943 GAGCCCTATGGAGAGAGGAGGCTCAGAGTCACTCTTTTCAAGAAAGCCAGCCCTAGGAAC 3002  
 DB 601 GAGCCCTATGGAGAGAGGAGGCTCAGAGTCACTCTTTTCAAGAAAGCCAGCCCTAGGAAC 660  
 QY 3003 CCAGAGAAAGAGTGTACTTAAAGCTGAATTTGTTTGTCTGTTGCCAGAGGTTGGTATT 3062  
 DB 661 CCAGAGAAAGAGTGTACTTAAAGCTGAATTTGTTTGTCTGTTGCCAGAGGTTGGTATT 720  
 QY 3063 GAGAGTAAACATTTATTTCTTTT 3088  
 DB 721 GAGAGTAAACATTTATTTCTTTT 746

## RESULT 15

BI416218 846 bp mRNA linear EST 14-AUG-2001  
 LOCUS 602989878F1 NCI\_CGAP\_Lu33 Mus musculus cDNA clone IMAGE:5145762 5',  
 DEFINITION mRNA sequence.

ACCESSION BI416218  
 VERSION BI416218  
 KEYWORDS GI:15177141  
 SOURCE EST.  
 ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 846)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@pdr@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1AM1359 row: 1 column: 19  
 High quality sequence start: 20  
 High quality sequence stop: 793.  
 Location/Qualifiers

## FEATURES

source  
 1..846  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Czech 1r"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5145762"  
 /label="pooled lung tumors"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NCI CGAP\_Lu33"  
 /note="Organ: lung; Vector: pRTT3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st  
 strand cDNA was prepared from mRNA obtained from pooled  
 lung tumors with a Not I - oligo(dT) primer [5'

## ORIGIN

TGTACCAATCTGAAGTGGAGGCGCCCTCTGTTTTTTTTTTT 3'.  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified pRT3 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bernaldo.

Query Match 23.8%; Score 739.2; DB 12; Length 846;  
 Best Local Similarity 97.1%; Pred. 1.6e-134;  
 Matches 816; Conservative 0; Mismatches 18; Indels 6; Gaps 6;

QY 1687 AGAAGTGAATGGAGAGCAACTGTGTGAGATGGTGTGACGAGGCTTATGTGACAGG 1746  
 DB 1 AGATCCGACAGAGTAGAGACACTGTGTGAGATGGTGTGACGAGGCTTATGTGACAGG 60  
 QY 1747 TGAATGTGCTCTTGTGACCAAAATATACCTAACCGCTGCCAAAATGGCTCTGTGAGCA 1806  
 DB 61 TGAATGTGCTCTTGTGACCAAAATATACCTAACCGCTGCCAAAATGGCTCTGTGAGCA 120  
 QY 1807 AGGCAACCCCTGAGTGTGATGGAGACGAGACTGTAGAGATGGCTCCGATGAGAAAACT 1866  
 DB 121 AGGCAACCCCTGAGTGTGATGGAGACGAGACTGTAGAGATGGCTCCGATGAGAAAACT 180  
 QY 1867 GTGACTGTGGCTGCGATCTTTACCAACAGGCTGCGTGTGTGTGACGAAATGCG 1926  
 DB 181 GTGACTGTGGCTGCGATCTTTACCAACAGGCTGCGTGTGTGTGACGAAATGCG 240  
 QY 1927 AGGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCAGGCTCGGGCCAGGAGCACTTGTGT 1986  
 DB 241 AGGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCAGGCTCGGGCCAGGAGCACTTGTGT 299  
 QY 1987 GGGCTGTGCTCATCTCTCTGAGTGGTGTCTGTGAGCTCATTTGCTTTCAGATGACA 2046  
 DB 300 GGGCTGTGCTCATCTCTCTGAGTGGTGTCTGTGAGCTCATTTGCTTTCAGATGACA 359  
 QY 2047 AAAATTTCAGTACTAGATCAACGATGTGAGACGCGCTTCCGGTCTGTGACAGCA 2106  
 DB 360 AAAATTTCAGTACTAGATCAACGATGTGAGACGCGCTTCCGGTCTGTGACAGCA 419  
 QY 2107 GCAGAGGCGAGTGGCTGCGGGGTGAGAGAGCTGAAGCTCAACGATCATGCCACCTT 2166  
 DB 420 GCAGAGGCGAGTGGCTGCGGGGTGAGAGAGCTGAAGCTCAACGATCATGCCACCTT 479  
 QY 2167 CCTTCATGATTTTCACTTGTGACTATGACATGCGCTTGTGAGCTGAGAGAGTGGT 2226  
 DB 480 CCTTCATGATTTTCACTTGTGACTATGACATGCGCTTGTGAGCTGAGAGAGTGGT 539  
 QY 2227 AGTACAGACCGTGTGTGCGGCCCATCTGCTGCTGTATGTCAACCATGTCTTCCCTG 2286  
 DB 540 AGTACAGACCGTGTGTGCGGCCCATCTGCTGCTGTATGTCAACCATGTCTTCCCTG 599  
 QY 2287 GC-AAGGCAATGGGTGACAGAGGCTGGGGGCACCAAAAGGAGGT-ACCGAGGCGT 2344  
 DB 600 GCAGAGGCGATCTGGGTGACAGAGGCTGGGGGCACCAAAAGGAGGTCAACCGAGGCT 659  
 QY 2345 GATCTTCAGAGAGGAGTGTGATCGTGTGATC-AAAGAGACCACTGTGAGAGCTCATGC 2403  
 DB 660 GATCTTCAGAGAGGAGTGTGATCGTGTGATC-AAAGAGACCACTGTGAGAGCTCATGC 719  
 QY 2404 GCGAGGAGTACCCCAAGATGTATGTGTGGTTTCTTCAGTGGGGGTGTGAGTCTT 2463  
 DB 720 GCGAG-AGATCACCCCAAGATGTATGTGTGGTTTCTTCAGTGGGGGTGTGAGTCTT 778  
 QY 2464 GCGAGGAGTACCTGTGTGCGCCCTTGTCAAGCGGAGAAAGATGGGGAATGTTCCAGG 2523  
 DB 779 G-CAGGGTGACTGTGTGCGCCCTTGTCAAGCGGAGCAAGATTTGGGAATGTTCCAGG 837

Search completed: September 30, 2004, 16:31:28  
 Job time : 7588 secs